PNAS RESEARCH ARTICLE | NEUROSCIENCE
PSYCHOLOGICAL AND COGNITIVE SCIENCES

Brain decoding of spontaneous thought: Predictive modeling of self-relevance and valence using personal narratives

Hong Ji Kim[a,b,c], Byeol Kim Lux[a,b,d], Eunjin Lee[a,b,c], Emily S. Finn[d], and Choong-Wan Woo[a,b,c,e,1]

Edited by Daniel Schacter, Harvard University, Cambridge, MA; received February 1, 2024; accepted February 20, 2024

The contents and dynamics of spontaneous thought are important factors for personality traits and mental health. However, assessing spontaneous thoughts is challenging due to their unconstrained nature, and directing participants' attention to report their thoughts may fundamentally alter them. Here, we aimed to decode two key content dimensions of spontaneous thought—self-relevance and valence—directly from brain activity. To train functional MRI-based predictive models, we used individually generated personal stories as stimuli in a story-reading task to mimic narrative-like spontaneous thoughts (n = 49). We then tested these models on multiple test datasets (total n = 199). The default mode, ventral attention, and frontoparietal networks played key roles in the predictions, with the anterior insula and midcingulate cortex contributing to self-relevance prediction and the left temporoparietal junction and dorsomedial prefrontal cortex contributing to valence prediction. Overall, this study presents brain models of internal thoughts and emotions, highlighting the potential for the brain decoding of spontaneous thought.

personal story | spontaneous thought | functional magnetic resonance imaging | brain decoding | affective neuroscience

Significance

Spontaneous thought provides valuable insights into our internal states and context, but assessing its contents and dynamics is challenging due to its unconstrained nature. We addressed this challenge by developing functional MRI-based predictive models for two crucial content dimensions (i.e., self-relevance and valence) of spontaneous thought. Using personalized narratives as stimuli, we evoked cognitive and affective responses resembling real-life experiences. Our models were able to predict the levels of self-relevance and valence ratings during story reading and resting state, contributing to brain-based daydream decoding. These results hold significant implications for understanding individual differences and assessing mental health, shedding light on the study of internal states and contexts that shape our subjective experiences.

Our mind never rests. Even during quiet periods or sleep, our mind spontaneously wanders from the past to the future and from one concept to another (1–3). Spontaneous thoughts may seem random, but they often involve topics that are emotionally charged, central to self-identity, and related to internal desires and goals (4, 5). The contents and dynamics of spontaneous thought are known to be important predictors of cognitive and affective traits (e.g., ruminative or internalizing cognitive styles) (2, 6–8) and disrupted brain processes, providing potential as cognitive and behavioral markers for mental and neurologic disorders, such as depression, anxiety, and Alzheimer's disease (9–11). However, the assessment of one's spontaneous thought is challenging, given that it occurs freely with minimal conscious constraints (12). In addition, the act of paying attention to spontaneous thought can change the nature of spontaneous thought itself, also known as the Heisenberg effect (13). For these reasons, measuring some aspects of spontaneous thought directly from brain activity, e.g., functional MRI (fMRI) signals, would be useful for the understanding of cognitive processes underlying spontaneous thoughts and the clinical application (14, 15). One previous study showed that the activation patterns within the medial orbitofrontal cortex (OFC) region from task-induced positive or negative affective states could classify positive vs. negative affective states during task-free rest (16). The current study took one step further by developing regression-based predictive models to decode the levels of two key affective dimensions of spontaneous thought—self-relevance and valence (6)—based on affective states elicited by personal narratives.

To effectively induce brain representations that resemble those of spontaneous thought, we chose to use narratives as stimuli. Recent studies have suggested that spontaneous thought is experienced in the form of images or words (17), particularly as deeply processed imagery and concepts such as narratives (18). Although narratives cannot capture all aspects of spontaneous thought, narratives share key elements with spontaneous thought, such as rich semantic information and their temporally unfolding nature (19). In addition, narratives have been successfully used in fMRI experiments to study semantic processing in the brain (20–25). Thus, narratives provide promising candidate materials for the study of brain representations of spontaneous thought.

However, the contents of spontaneous thought have an important characteristic that narratives created by others (e.g., experimenters) are lacking, which is that they are often very personally relevant (in other words, have high self-relevance) (7, 26, 27). The significance of personally relevant topics in spontaneous thought has been empirically demonstrated in multiple previous studies (7, 28–33). For example, Andrews-Hanna et al. reported that participants rated their spontaneous thoughts as having a high level of personal significance across two datasets (28), while Baird et al. found that 66% of the off-task thoughts were related to self (29). This motivated us to use personal

Author contributions: H.J.K. and C.-W.W. designed research; H.J.K., B.K.L., E.L., and C.-W.W. collected data; H.J.K. and C.-W.W. analyzed data; and H.J.K., E.S.F., and C.-W.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: waniwoo@skku.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2401959121/-/DCSupplemental.

Published March 28, 2024.

narratives in our experiment. It is well known that spontaneous thought is most commonly about one's personal life, such as current personal concerns, past memories, and future plans (5, 7, 33–35), suggesting that self-relevant thought contents are major building blocks of spontaneous thought. For example, a recent paper proposed that episodic memory serves as a foundation of spontaneous thought and provides a scaffolding for semantic memory to generate thought contents (36). In addition, self-referential information recruits brain systems distinct from those for non-self-referential information (37–40), underscoring the importance of using self-relevant stimuli to study brain representations of spontaneous thought. Thus, we hypothesized that personal narratives would be able to induce brain representations close to those of spontaneous thought. Here, we performed one-on-one interviews with participants to create individually unique stimuli based on personal narratives, which were used in the fMRI experiment to elicit self-relevant thoughts and emotions.

Among multiple dimensions of spontaneous thought (6, 7, 17), here we particularly focused on two content dimensions—self-relevance and valence. We chose these two dimensions mainly because dimensionality reduction analyses conducted in previous studies showed that self-relevance and valence were among the most central dimensions that can serve as summaries of other content dimensions (6, 7). In addition, these two variables are likely to be among the fundamental dimensions of human cognition and emotions given that self-relevance and valence convey information crucial for survival, such as what to avoid (i.e., negative valence, high self-relevance), what to approach (i.e., positive valence, high self-relevance), or what to ignore (i.e., low self-relevance). In terms of the brain systems, self-relevance and valence can be linked to the brain networks related to valuation (i.e., is it good or bad?) and context-dependent salience detection (i.e., is it relevant to me?), such as the default mode, limbic, and ventral attention networks (41–43). Note that, however, these two dimensions are only a small fraction of components that constitute spontaneous thoughts, and therefore, consider this study as a step toward the decoding of the rich content of spontaneous thoughts.

As shown in Fig. 1*A*, the major research goals of the current study include 1) developing fMRI multivariate pattern-based predictive models of self-relevance and valence using data from the story-reading task, 2) comparing and interpreting the newly developed predictive models of self-relevance and valence, and 3) testing the predictive models on the resting-state fMRI data with and without intermittent thought-sampling probes. To this end, we conducted an fMRI experiment (n = 49) while participants underwent the story-reading and thought-sampling tasks. In the story-reading task, participants were asked to read their own stories or stories made by others to induce a wide range of levels of self-relevance and valence. In the thought-sampling task, participants were asked to think freely and intermittently report a few words that represented their current thoughts. After fMRI scans, participants provided self-relevance and valence ratings for the stories and words from the story-reading and thought-sampling tasks. With the fMRI data from the story-reading task, we developed fMRI multivariate pattern-based decoding models of self-relevance and valence that showed significant predictions in the leave-one-subject-out cross-validation (LOSO-CV). We then identified important contributors to the prediction of both models using the virtual lesion and isolation analysis methods. Finally, we applied these models to decode self-relevance and valence scores during the thought-sampling task (n = 49) and resting state (n = 90 and 60).

Results

Experimental Overview and Post-scan Survey. Fig. 1*B* shows the experimental design of the current study, which we briefly describe here (for the details of the experimental procedure, please see *Materials and Methods* and *SI Appendix*, Supplementary Methods). On day 1, we conducted a one-on-one interview with participants to create personal stories to use as stimuli in the fMRI experiment. On day 2, participants underwent the story-reading and thought-sampling tasks in the scanner. During the story-reading task, participants read four "personal" stories, which were created for each participant, and six "common" stories, which were the same across participants. While reading the stories, participants were intermittently asked to provide their valence ratings (i.e., three times per story). In the thought-sampling task, we asked participants to think freely and verbally report what was in their mind with a few words intermittently [every 50.7 ± 5.6 (mean ± SD) s]. After the scan, we conducted a post-scan survey on words and stories (Fig. 1*C*). For the word survey, participants rated the words they generated during the thought-sampling task using a multidimensional content scale (see *SI Appendix*, Supplementary Methods for details), and for the story survey, participants read the stories again and rated their perceived levels of self-relevance and valence using continuous ratings (see *SI Appendix*, Fig. S1 for example and group average ratings). To ensure that the post-scan survey results reflected the in-scanner experience, we compared the intermittent valence ratings from the in-scanner story-reading task with the valence ratings from the post-scan survey. As shown in Fig. 1*D*, the in-scanner vs. post-scan valence ratings were highly correlated (mean r = 0.844, z = 51.71, P < 2.220e-16, two-tailed, bootstrap tests with 10,000 iterations), suggesting that the post-scan survey ratings reflected the in-scanner experience well.

Developing Predictive Models of Self-relevance and Valence. To achieve the first research goal, i.e., developing fMRI-based predictive models of self-relevance and valence (Fig. 1*A*), we trained fMRI multivariate pattern-based predictive models using the fMRI data from the story-reading task. To effectively disentangle fMRI patterns for self-relevance and valence, we concatenated and quantized the data based on quintiles of the two dimensions, constructing 25 averaged images per participant (i.e., 5 levels of self-relevance × 5 levels of valence; for details of the data quantization and distribution of ratings, see *SI Appendix*, Figs. S2 and S3). We then trained predictive models of self-relevance and valence using principal component regression (PCR) (44) with LOSO-CV and random-split cross-validation (RS-CV) (45, 46). The models showed significant cross-validated prediction performances—prediction–outcome correlations for self-relevance, with LOSO-CV, mean r = 0.322, z = 9.204, P < 2.220e-16, mean squared error (mse) = 0.148, two-tailed, bootstrap tests with 10,000 iterations, with RS-CV, mean r = 0.332, mse = 0.144 (Fig. 2*A*); for valence, with LOSO-CV, mean r = 0.205, z = 6.235, P = 4.511e-10, mse = 0.454, two-tailed, bootstrap tests with 10,000 iterations, with RS-CV, mean r = 0.179, mse = 0.458 (Fig. 2*B*). Permutation tests with 1,000 iterations and LOSO-CV also provided significant performances against null models (both P = 0.0010; *SI Appendix*, Fig. S4). Interestingly, an additional analysis correlating model performances with vividness ratings across subjects showed a significant positive correlation for the valence model, r = 0.370, P = 0.0090 (*SI Appendix*, Fig. S5), suggesting that poorer prediction performance for some individuals might arise from reduced engagement with the narratives.

As shown in Fig. 2 *A* and *B*, the two models showed a weak correlation between the unthresholded whole-brain patterns of

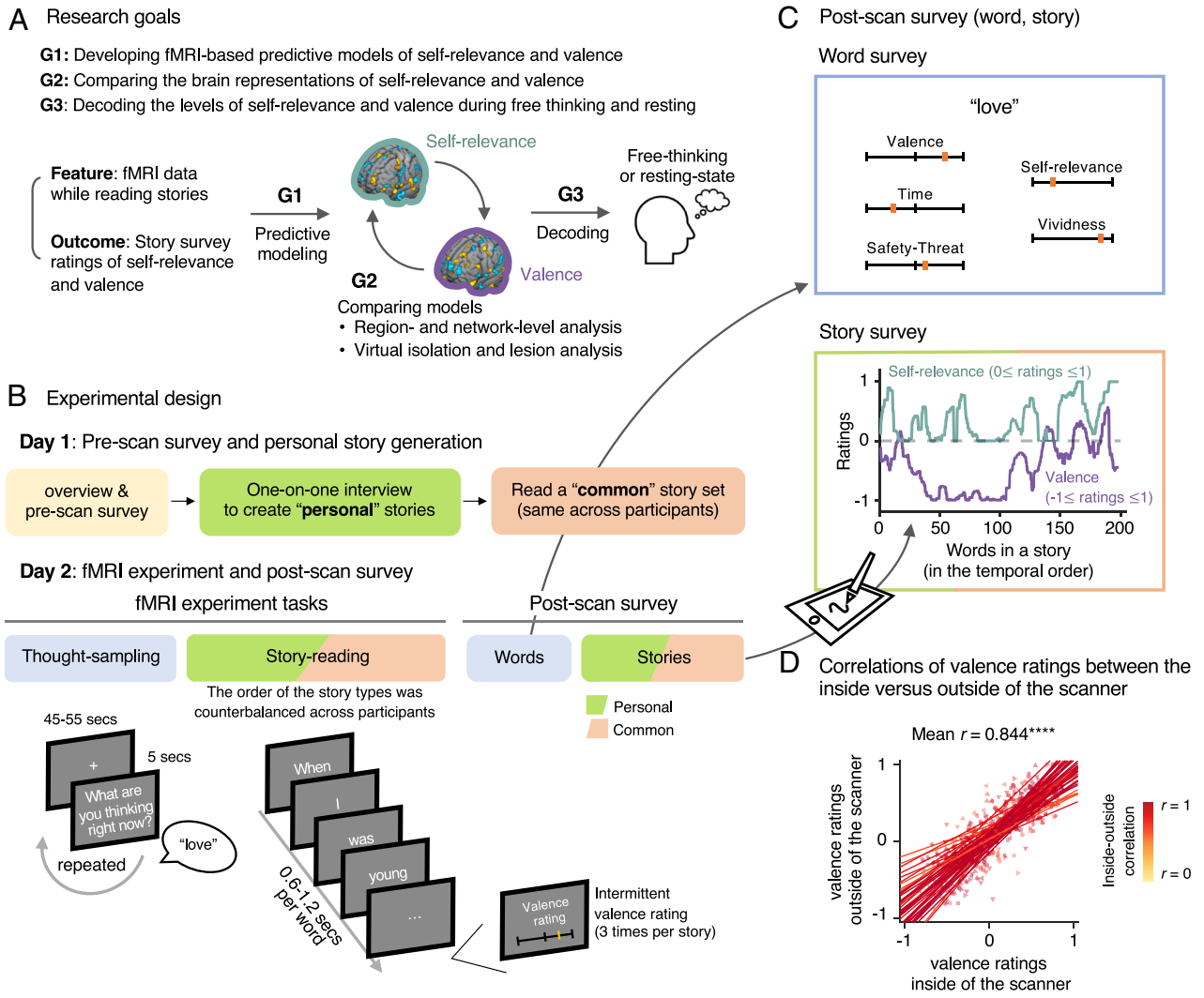

Fig. 1. Research goals and experimental design. (*A*) This study has three research goals, each corresponding to different analysis steps. The first goal is to build predictive models, the second goal is to interpret the model features that contribute most to prediction, and the third goal is to test the models on independent data. (*B*) The experiment was conducted over 2 d separated by an interval of approximately 1 wk (mean = 7.3 d). On day 1, we conducted a pre-scan survey and one-on-one interviews to create personal story stimuli. On day 2, we conducted an fMRI experiment that consisted of five story-reading runs and two thought-sampling runs. During the story-reading runs, we asked participants to intermittently rate the current emotional valence (three times per story). For details of the experimental procedure, please see *Materials and Methods* and *SI Appendix*, Supplementary Methods. (*C*) After the scan, participants underwent the post-scan survey for thought-sampling responses (i.e., words or phrases) and stories. For each thought-sampling response, we asked participants to rate the response on five content dimensions, including self-relevance, valence, and other dimensions. For the story survey, we asked participants to read the stories again while continuously rating them on the content dimensions of self-relevance and valence using a tablet pen. The plot shows example story ratings, in which the green line indicates self-relevance ratings, while the purple line indicates valence ratings. (*D*) The valence ratings between the inside and the outside of the scanner showed high within-individual correlations, mean $r = 0.844$, $P < 2.220e\text{-}16$, two-tailed, bootstrap tests with 10,000 iterations.

predictive weights, $r = 0.120$, while the mPFC, which is known to be important both for self-referential and valence information processing (47–50), showed a weak, but larger correlation ($r = 0.269$) than the whole brain, suggesting the existence of overlapping representations between self-relevance and valence within the mPFC. For example, both self-relevance and valence models thresholded at $P < 0.001$ (two-tailed, bootstrap test) showed negative weights within the dorsomedial prefrontal cortex [dmPFC, Brodmann area (BA) 9] and the subgenual anterior cingulate cortex (sgACC, BA25) and positive weights in the ventromedial prefrontal cortex (vmPFC, BA11). However, given that these are the uncorrected results, we conducted further analyses on the network-level and voxel-level importance, the results of which we describe in the next section.

We also conducted additional analyses on the self-relevance and valence models to further examine their validity. First, we tested whether the self-relevance model could classify the personal versus common stories. As Fig. 2*C* shows, the self-relevance ratings were significantly higher in the personal stories [0.75 ± 0.16 (mean ± SD)] than in the common stories (0.42 ± 0.10), $t_{48} = 15.86$, $P < 2.220e\text{-}16$, two-tailed, paired *t*-test, consistent with our assumption that the self-generated personal stories would be more self-relevant than common stories generated by others. The cross-validated responses of the self-relevance model were also significantly higher for the personal stories than the common stories, $t_{48} = 10.18$, $P < 2.220e\text{-}16$, two-tailed, paired *t*-test. The forced-choice classification accuracy with LOSO-CV was 93.8%, $P = 6.980e\text{-}11$, two-tailed, binomial test. These results suggest that reading personal and common stories induces different brain representations, and the differences can be captured by our self-relevance model.

Second, to examine the validity of the valence model, we selected the top 20 positive and negative words from the six common story sets based either on the TR-by-TR actual valence ratings or the

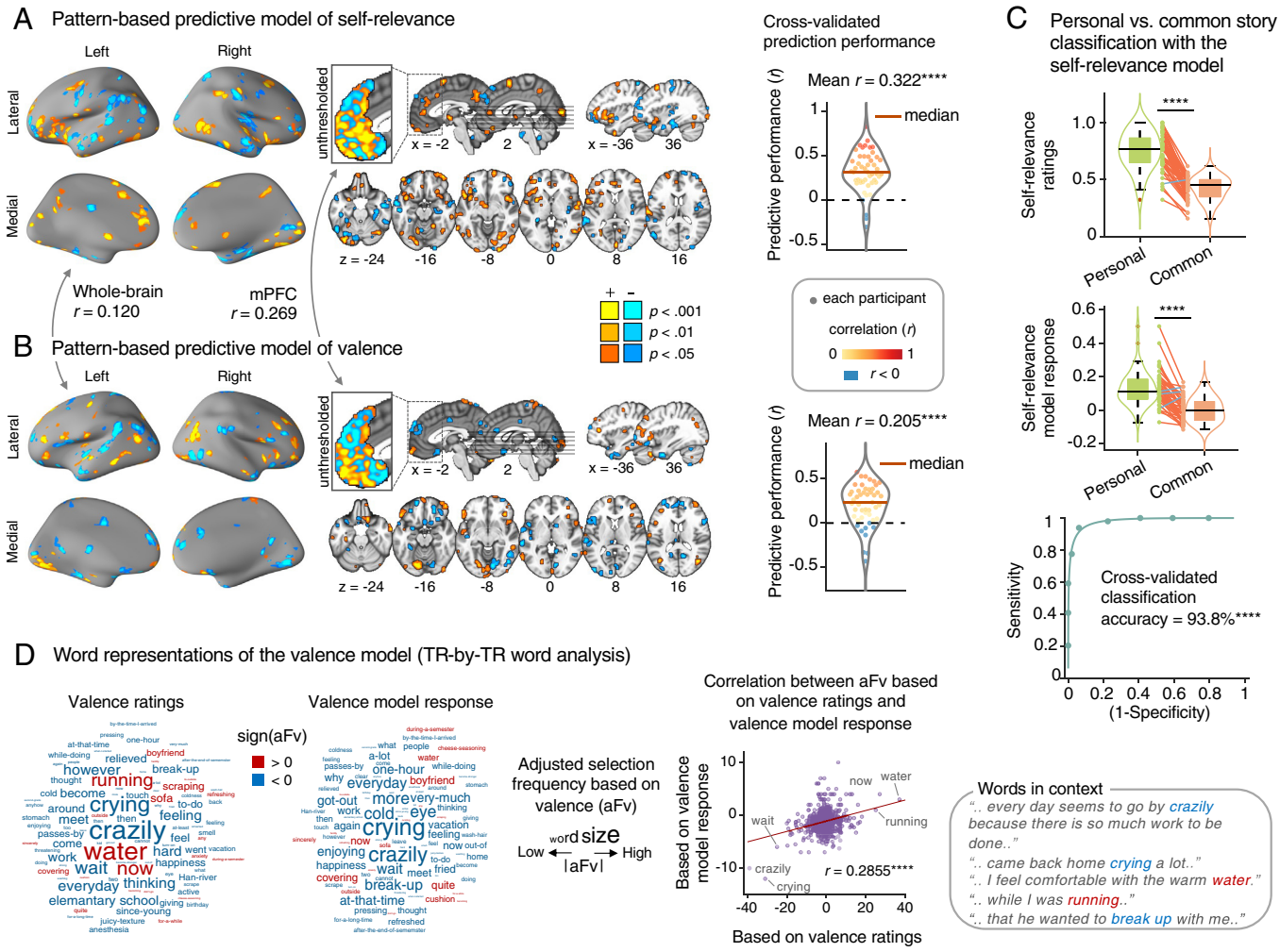

Fig. 2. Multivariate pattern-based predictive models of self-relevance and valence. (*A*) Multivariate pattern-based predictive model of self-relevance. The brain map shows the predictive weights (positive in warm colors, negative in cool colors) that reliably contributed to the prediction of self-relevance based on bootstrap tests (thresholded at uncorrected $P < 0.001$, two-tailed). We thresholded the map for the purpose of display and interpretation; all weights were used in the prediction. We also pruned the map using two more liberal thresholds, uncorrected $P < 0.01$ and $P < 0.05$, two-tailed, to show the extent of activation clusters. The *Inset* shows the unthresholded weight patterns within the medial prefrontal cortex (mPFC). The violin plot on the right shows the LOSO-CV model performance, $n = 49$, mean $r = 0.322$, $P < 2.220e\text{-}16$, bootstrap test, two-tailed. Each dot indicates the predictive performance of each participant, and the thick red line indicates the median. $^{****}P < 0.0001$. (*B*) Multivariate pattern-based predictive models of valence. (*C*) (*Top*) The personal stories showed a higher level of self-relevance ratings compared to the common stories, $t_{48} = 15.86$, $P < 2.2204e\text{-}16$, two-tailed, paired *t*-test. orange line: higher scores in the personal stories, blue line: higher scores in the common stories. (*Middle* and *Bottom*) The personal stories showed a higher level of the self-relevance model response than the common stories, $t_{48} = 10.18$, $P < 2.2204e\text{-}16$, and the forced-choice classification accuracy was 93.8%, $P = 6.980e\text{-}11$, two-tailed, binomial test. (*D*) Word representations of the valence model in TR-by-TR word analysis. We used the adjusted selection frequency based on valence (aFv; see *SI Appendix*, Supplementary Methods) to examine the relative importance of words based on valence ratings or valence model responses. The wordclouds show the relative aFv scores for the top 100 words selected based on the absolute values of actual valence ratings (*Left*) and the model responses on those words (*Middle*). The word size represents the absolute magnitude of aFv. The scatter plot shows the relationship between the aFv scores based on the valence ratings (*x* axis) and the valence model responses (*y* axis) from the whole word set. The "words-in-context" box shows the context in which some top words were used.

TR-by-TR valence model responses. We then compared the selection frequency between the ratings vs. model responses after adjusting the frequency value with the overall word frequency, which we named adjusted selection frequency based on valence (aFv; for details, see *SI Appendix*, Supplementary Methods). As shown in Fig. 2*D*, the aFv values based on the actual valence ratings showed a significant correlation with the aFv values based on the valence model, $r = 0.2855$, $P = 1.110e\text{-}16$. For example, the words "crazily," "crying," and "wait" had low aFv values based on ratings and model response, whereas "water," "now," and "running" showed high aFv values in both measures. Fig. 2*D* also provides the contexts in which these words were used. We applied the same approach to the self-relevance model as well, shown in *SI Appendix*, Fig. S6.

We also conducted univariate general linear model (GLM) analyses for the comparisons (*SI Appendix*, Fig. S7). Although the univariate maps showed activation patterns distinct from the predictive models (whole-brain pattern similarity between the GLM and relevant predictive model maps, $r = 0.302$ for valence and $r = 0.226$ for self-relevance, *SI Appendix*, Fig. S7 *A* and *B*), there were also some consistent findings. For example, the contrast map for the personal vs. common stories (*SI Appendix*, Fig. S7*C*) and the parametric modulation map of self-relevance ratings showed strong activations within the anterior midcingulate cortex (aMCC) and anterior insula (aINS), which were consistent with the multivariate pattern-based predictive model of self-relevance ($r = 0.142$ between the contrast map and self-relevance predictive model). In addition, the parametric modulation map of valence ratings showed positive vmPFC activation, consistent with the multivariate pattern-based valence model.

Network- and Region-level Importance for Predictive Models. To achieve the second research goal, i.e., comparing brain representations of self-relevance and valence (Fig. 1*A*), we further

examined the importance of various features for the self-relevance and valence models (51). Specifically, we examined network- and region-level importance with virtual isolation analysis (i.e., calculating the prediction–outcome correlation based on a single large-scale network, region, or searchlight at a time) and virtual lesion analysis (51) (i.e., calculating the changes in the prediction–outcome correlation after removing one network, region, or searchlight at a time). As shown in Fig. 3*A*, the virtual isolation analysis using large-scale networks and some regions of interest (ROIs) showed that the default mode, ventral attention, and frontoparietal networks had significant prediction performances for both self-relevance and valence (bootstrap test with 10,000 iterations; for details, see *SI Appendix*, Table S1). For the brain maps of the large-scale networks and ROIs, please see *SI Appendix*, Fig. S8. When we tested the mPFC region separately, it also showed significant prediction for both self-relevance and valence. The visual network was, however, important only for predicting self-relevance, while the limbic network was important only for predicting valence. The virtual lesion analysis using large-scale networks and ROIs also showed that the default mode network (DMN) was important for predicting both self-relevance and valence (*SI Appendix*, Fig. S9*A* and Table S1), while the visual, ventral attention, and frontoparietal networks were important only for predicting self-relevance, and the mPFC was important only for predicting valence.

The Fig. 3*B* and *SI Appendix*, Fig. S9*B* present the searchlight analysis results for the virtual isolation and virtual lesion analyses, respectively. The searchlight-based virtual isolation results show that the aMCC, aINS, and visual areas were important for the prediction of self-relevance, and for the valence model, the dmPFC, temporoparietal junction (TPJ), temporal pole (TP), and other regions were important predictors. Fig. 3*C* shows the conjunction map between the importance maps of self-relevance and valence models, and the supplementary motor area (SMA), superior temporal gyrus (STG), and inferior frontal gyrus (IFG) appeared to be important for both models.

Decoding the Levels of Self-relevance and Valence during Free-thinking and Resting. To achieve the third research goal, i.e., decoding the levels of self-relevance and valence during free-thinking and resting (Fig. 1*A*), we first tested the models on the fMRI data from the thought-sampling task, in which we asked participants to report what they were thinking every 50 s with jitters [interval = 50.7 ± 5.6 (mean ± SD) s; see *SI Appendix*, Supplementary Methods for the details of the task]. Given that the verbal responses to thought sampling are most likely to be based on the thought contents just before the reporting onset and considering the hemodynamic delay, as shown in Fig. 4*A*, our time of interest was a 10-TR (a total of 4.6 s) time window around the reporting onset. We evaluated the prediction performance with the prediction–outcome correlations with LOSO-CV using the moving-window approach based on the data convolved with the temporal Gaussian kernel (FWHM = 10 TRs). As in Fig. 4 *B*, *Left* panel, both predictions of self-relevance and valence showed the best prediction performance at the time-of-interest period. When we examined the prediction performances of time bins of 10 TRs (Fig. 4 *B*, *Right* panel), both models showed weak but significant correlations for the corresponding ratings at the time-of-interest bin (for the self-relevance model predicting self-relevance ratings, mean r = 0.0518, P = 0.0141, one-tailed, bootstrap test with 10,000 iterations; for the valence model predicting valence ratings, mean r = 0.0495, P = 0.0095). For the exact r and p values of each time bin, please see *SI Appendix*, Table S2. To compare our model performance with other predefined models, we also tested nine a priori maps from previous studies in the same manner. The a priori maps included two maps for cognitive components of self-generated thought (52), six meta-analytic maps of aversion, episodic, default, self, emotion, and semantic from a previous study (53), and the picture-induced negative emotion signature (PINES) (54) (Fig. 4*C*). The result showed that only the models from this study were predictive of the self-relevance and valence ratings (see *SI Appendix*, Fig. S10*A* for predictive performances of a priori maps in all time bins). We also examined which large-scale

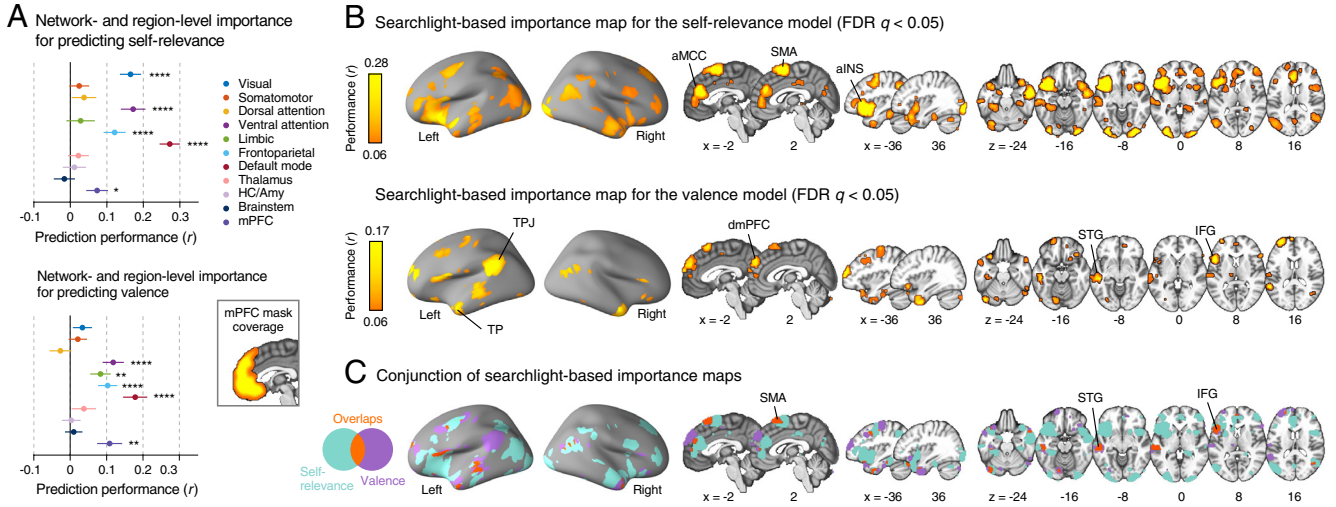

Fig. 3. Important features of the self-relevance and valence models (virtual isolation analysis). We examined which features were important for predicting self-relevance and valence using the virtual isolation analysis, which calculated prediction performance (i.e., prediction–outcome correlation) based on a single large-scale network, region, or searchlight at a time. (*A*) The virtual isolation analysis results for the self-relevance model (*Top*) and the valence model (*Bottom*) with the large-scale networks and some ROIs. Each colored dot represents the prediction–outcome correlations for each network or region with bootstrap tests with 10,000 iterations. The error bars represent the SD of the sampling distribution. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. (*B*) Searchlight-based virtual isolation analysis results for the self-relevance model (*Top*) and the valence model (*Bottom*). The maps were thresholded at FDR $q < 0.05$, two-tailed, bootstrap tests. (*C*) Conjunction of the searchlight-based importance maps for the self-relevance and valence models. aINS, anterior insula; aMCC, anterior mid-cingulate cortex; dmPFC, dorsomedial prefrontal cortex; IFG, inferior frontal gyrus; SMA, supplementary motor area; STG, superior temporal gyrus; TP, temporal pole; TPJ, temporoparietal junction. For the brain maps of the large-scale networks and ROIs, please see *SI Appendix*, Fig. S8.

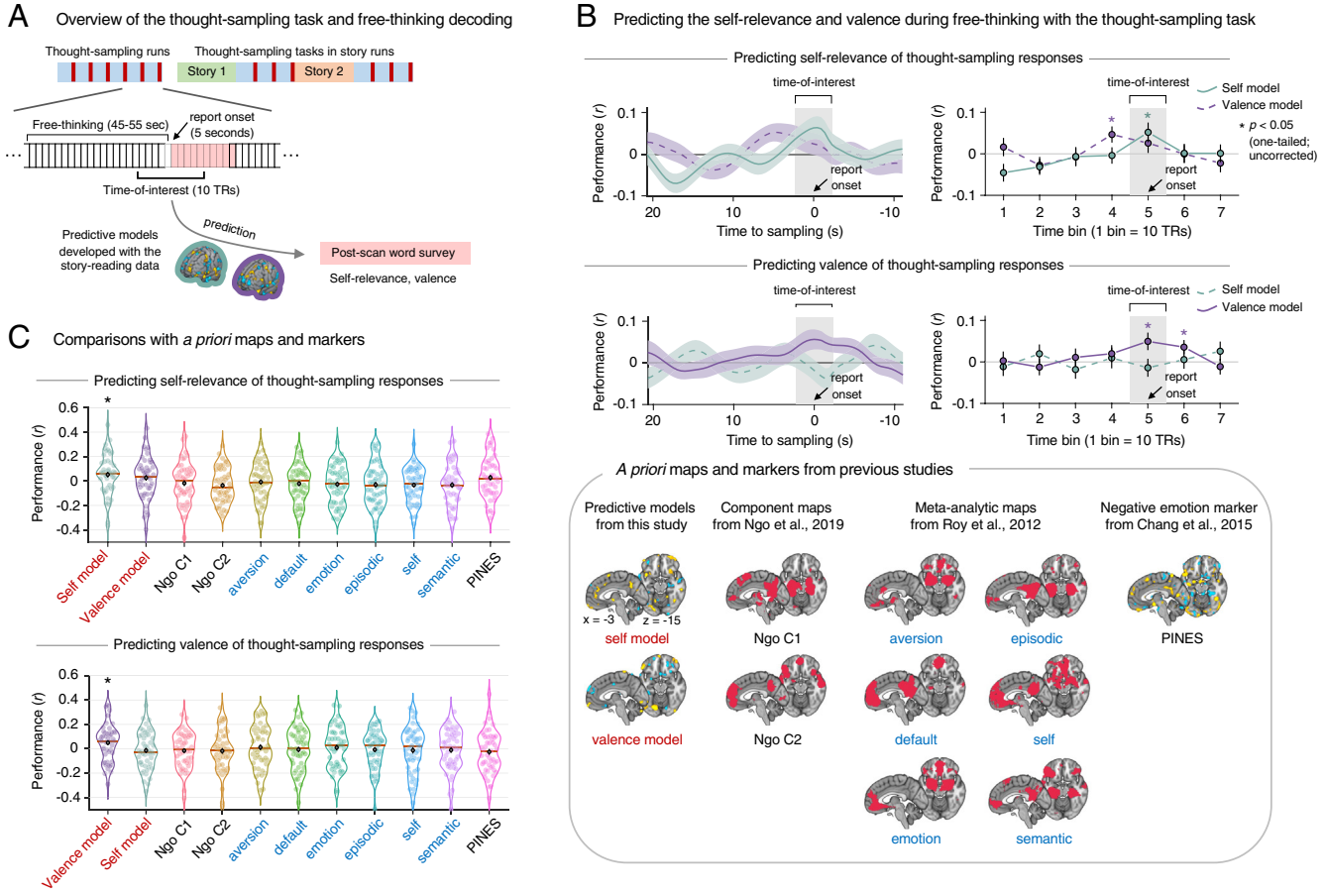

Fig. 4. Decoding self-relevance and valence during free-thinking. (*A*) During the thought-sampling task, participants were instructed to think freely for around 50 s and then verbally report what they were thinking in words or phrases. One thought-sampling run included a total of six trials, and each story-reading was followed by three trials of thought-sampling. We applied the predictive models of self-relevance and valence to the fMRI data from the free-thinking period to predict the self-relevance and valence ratings collected from the post-scan survey. (*B*) Model performances (*Top*: predicting self-relevance; *Bottom*: predicting valence) measured by the prediction–outcome correlations. The plots on the left show the prediction performance using a moving-window approach based on the data convolved with a temporal Gaussian kernel (FWHM = 10 TRs). The plots on the right show the prediction performances for the time bins of 10 TRs. The predictive performances were calculated as prediction–outcome correlations for each time bin, with one-tailed test and bootstrap tests with 10,000 iterations. $^{*}P < 0.05$. Shading and error bars indicate the SD of performances across all participants. (*C*) The prediction performances of the self-relevance and valence models from the current study and nine a priori maps (the a priori maps are shown on the right). The a priori maps included two component maps of self-generated thought (52), six meta-analytic maps (53), and the PINES (54).

networks and ROIs were important for these predictions (at the time-of-interest bin) with the virtual isolation analysis and found that the DMN was the only predictor important for both self-relevance and valence models (*SI Appendix*, Fig. S11, and for the exact r and P values of each time bin, please see *SI Appendix*, Table S3).

Finally, we tested the self-relevance and valence models on the resting-state fMRI data from two independent datasets (n = 90 and 60), in which, at the end of the resting scan, participants reported the levels of self-relevance and valence for the thoughts they had during the resting scan (post-resting survey; Fig. 5*A* and *SI Appendix*, Fig. S12). Since the answers to the post-resting survey were likely to be based on participants' thoughts near the end of the scan, we evaluated the model performance using averaged data from the end of the scan (see *SI Appendix*, Fig. S13 for the model performance over the entire time series of the run). As shown in Fig. 5*B*, for the first independent dataset (n = 90), the self-relevance and valence models showed significant predictions for the self-relevance and valence ratings around the same temporal window, which was 27th to 33rd TRs (for self-relevance) and 26th to 37th TRs (for valence) from the end of the scan (the gray boxes in Fig. 5*B*, which indicated the time points with $r > 0$ and $P < 0.05$, one-tailed, one-sample t-test). The best prediction was observed when the data of the last 31 TRs (i.e., 14.3 s) were averaged, and the prediction performances were r = 0.189, P = 0.037 for the self-relevance model and r = 0.300, P = 0.002 for the valence model (one-tailed, scatter plots in Fig. 5*B*). Similar to what we did for the free-thinking decoding, we examined which large-scale networks and ROIs were important for these predictions (for the last 31 TRs) with the virtual isolation analysis. For self-relevance, only the DMN was significant, while for valence, the ventral attention, limbic networks, and brainstem were significant (Fig. 5*C*, also see *SI Appendix*, Fig. S10*B* for predictive performances of a priori maps for the resting-state decoding). We additionally tested our models on the second independent resting-state dataset (n = 60; *SI Appendix*, Fig. S12*A*). Taking the last 31 TR-sized time window as a predefined hypothesis, the valence model showed a significant prediction performance (r = 0.320, P = 0.0063), while the self-relevance model did not demonstrate significant prediction (r = −0.094, P = 0.2375). However, further exploratory analyses on varying window sizes (*SI Appendix*, Fig. S12*B*) revealed that an average of the last 10 TRs (equivalent to 4.6 s) yielded a significant predictive performance (r = 0.218, P = 0.0470) in predicting self-relevance.

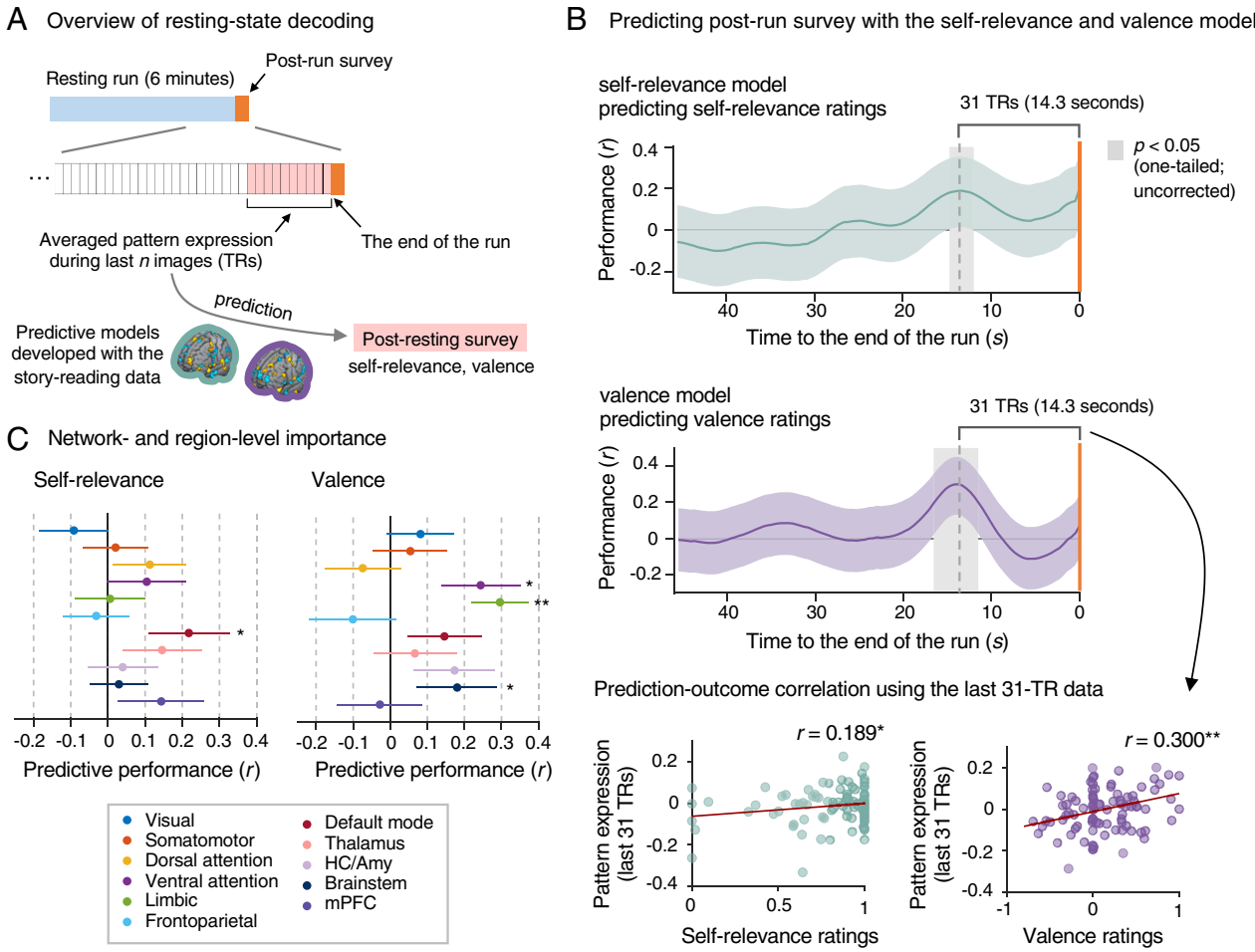

Fig. 5. Decoding self-relevance and valence during rest. (*A*) We tested our models on data from a 6-min resting-state run from an independent dataset (n = 90) to predict self-relevance and valence ratings from a post-run survey on spontaneous thoughts during resting. (*B*) Prediction performance was calculated as the prediction–outcome correlation based on the averaged data from the end of the scan. The gray boxes indicate the time period where each model showed significant positive predictions (i.e., $P < 0.05$ and $r > 0$). The scatter plots show the relationship between the ratings and pattern expression values when the last 31 TRs (14.3 s) were averaged. Each dot indicates each participant. $^{*}P < 0.05$ and $^{**}P < 0.01$, one-tailed, one-sample t-test. (*C*) Network- and region-level decoding performance of the last 31 TRs. Each colored dot represents the mean prediction–outcome correlations for each network or region with bootstrap tests with 10,000 iterations. The error bars represent the SD of the sampling distribution, $^{*}P < 0.05$ and $^{**}P < 0.01$, one-tailed, bootstrap test.

Discussion

In this study, we developed multivariate pattern-based predictive models of self-relevance and valence that can be used to decode affective dimensions of spontaneous thoughts. For this, we conducted an fMRI experiment using a narrative-reading task, in which we showed personal stories created through one-on-one online interviews with participants and common stories created by others. The self-relevance and valence models could decode the levels of these two content dimensions of spontaneous thought during rest across three datasets. The main innovations and findings of this study can be summarized as the following: 1) we used personal stories as stimuli to evoke thoughts and emotions that resemble spontaneous thoughts; 2) we identified brain systems important for decoding self-relevance and valence using virtual isolation and lesion analyses, such as the mPFC, aINS, and TPJ; and 3) we were able to decode the level of self-relevance and valence during rest with or without thought sampling using the predictive models.

First, we used personal narratives as experimental stimuli to induce cognitive and emotional states similar to spontaneous thoughts. We used self-generated personal stories as stimuli for the following reasons: 1) self-relevant thoughts, such as current personal concerns, past memories, and future plans, are known to feature prominently in spontaneous thought (7, 26, 27); 2) recent studies suggested that spontaneous thoughts are experienced in the form of deeply processed imagery and concepts, such as narratives (18); and 3) brain structures important for spontaneous thought and internal narrative construction are largely overlapping (55–57). Therefore, personal stories based on participants' past experiences and related emotions should share many characteristics with spontaneous thoughts. In addition, following recent efforts in the neuroimaging field to use naturalistic stimuli, such as movies (24, 58, 59), we created individually unique stories through one-on-one interviews with participants to make the experimental stimuli as natural as possible. The common stories used for all participants were also created by pilot participants through the same one-on-one interview procedure. Therefore, we believe our story stimuli were able to induce natural and vivid thoughts and emotions that resemble spontaneous thoughts during the scan. Using these stories as a bridge between the experiment and naturally experiencing spontaneous thoughts, we developed fMRI-based predictive models that could decode not only the contents of current reading materials but also the phenomenological qualities of individuals' spontaneous thoughts (15).

Second, we identified brain regions and networks important for decoding levels of self-relevance and valence. Using virtual isolation and virtual lesion methods (51), we found some converging

evidence that the DMN, including the mPFC, plays an important role in predicting both self-relevance and valence. Particularly, for decoding levels of self-relevance and valence during free-thinking, the DMN was the only significant predictor for both self-relevance and valence (*SI Appendix*, Fig. S10). This is consistent with previous literature suggesting that the DMN is important for stimulus-independent, task-unrelated thought (14), processing details of ongoing cognition (17), memory (60, 61), affective content of spontaneous thought (16), and self (62, 63). There were also brain networks and regions that were uniquely important for each model. For example, the visual network (Fig. 3*A*), aINS, and aMCC (Fig. 3*B*) were among the important features for decoding self-relevance, but these were not important for valence. The importance of the aINS and aMCC for self-relevance can be understood in the context of the salience network (or the ventral attention network), considering that the aINS and aMCC are among the key hubs of the network. The salience network is known to be important for context-dependent salience detection (42, 43), and it is likely that when there is no external task (such as reading autobiographical stories without external tasks), the self can become the most important context, which is generated from within. For example, in our task, the primary concern or ongoing implicit task could become identifying content relevant to oneself. This could be the case for many other naturalistic situations where we do not have external tasks. In addition, the limbic network (Fig. 3*A*) and the left TPJ and dmPFC (Fig. 3*B*) were among the important features for decoding valence, but these were not important for self-relevance. The importance of the limbic network in emotional valence is also consistent with previous studies reporting that the OFC and TP, which are the main components of the limbic network, were important for valence processing (50, 64). We also found that the left TPJ, along with the dmPFC (Fig. 3*B*), were important for valence. This finding was rather unexpected, but the left TPJ and dmPFC are the major constituents of DMN's dorsal medial subsystem (65). This subsystem has been suggested to play an important role in a high-level "mind's mind" form of imagination, such as reflective thinking in the verbal form (55). Thus, we can speculate that feeling positive or negative emotions from reading autobiographical stories requires more of the mind's mind form of imagination, such as mentalizing (66).

Third, with our self-relevance and valence models, we were able to decode the respective content dimension scores during free-thinking and resting. Different from recent efforts to decode semantic features directly from brain activity (25, 67, 68), we targeted the affective dimensions of thought, which provide information that is complementary yet somewhat independent of its semantic dimension. The affective dimension becomes particularly important when it comes to concepts of personal relevance (e.g., those related to autobiographical memory or personal background). For instance, even if two individuals contemplate the same concept (e.g., father), their personal meanings can diverge dramatically (e.g., caring father vs. abusive father), which can be reflected in the affective dimension. Our study focused on the affective dimensions of spontaneous thought to capture the idiosyncrasies with which individuals perceive and relate to semantic concepts. There was a previous study that also tried to classify positive versus negatively valenced task-free thoughts based on task-induced brain activation patterns, especially focusing on the medial orbitofrontal cortex (16). Our study has a similar motivation and approach in that we also used task-induced brain activation patterns to develop decoders for the affective content dimensions of spontaneous thought. However, our approach has extended the previous study in multiple aspects. Different from building a classification model based on a local region's activation patterns, we developed regression-based predictive models using whole-brain activation patterns. In addition, our modeling targeted to predict the levels of self-relevance as well as valence. Our self-relevance and valence models showed weak but significant prediction performances across three different datasets with two different task contexts—i.e., thought sampling and resting. This represents a powerful combination of task-based data (i.e., personal and common stories) with annotated rest in an effort to decode the contents of the mind during rest using a known ground truth (69).

Given that the contents and dynamics of spontaneous thought could provide rich information about individuals' mental and brain health, the ability to decode some aspects of spontaneous thought directly from neuroimaging data would be useful. In this study, we focused on two content dimensions of spontaneous thought—self-relevance and valence, which are among the important predictors of depression and negative affectivity scores in subclinical populations (6, 7). Though further studies, including clinical ones, are needed to identify which content dimensions are most useful for predicting and promoting mental health and well-being, this study showed the potential to use fMRI data to extract certain information about spontaneous thought. Importantly, we demonstrated that resting-state fMRI could be used to decode aspects of spontaneous thought, opening a broad avenue for using resting-state data involving no or minimal tasks (e.g., introspection) to obtain rich information about one's internal cognition. This is important because it is often difficult (or impossible) to administer task-based fMRI tests to patients. In addition, our study further implies that we should reconceptualize the notion of the resting state. As we wrote in the introduction, our mind (and brain) never rests. Our mind keeps wandering, and our brain keeps being activated spontaneously. Therefore, the resting state should be reconceptualized as spontaneous cognition or spontaneous activity conditions. With this perspective, we will be able to develop tasks that are based on resting but have minimal task components, which then can be used to provide richer and clinically more useful information.

There are some limitations in this study. First, we used two different types of stories—personal and common stories—to increase the variance of the level of self-relevance, the orthogonality between self-relevance and valence, and the comparability across participants. However, it is also possible that they are qualitatively different and thus might not be comparable between the two conditions. We made multiple efforts to resolve this issue. Given that we expected personal stories to be more familiar than common stories, we had participants read all common stories on day 1. We also gave quizzes on the common stories twice—first, on day 1 after reading the stories, and second, on day 2 right before the fMRI experiment (see *SI Appendix*, *Supplementary Methods* for the quiz scores). In addition, we collected ratings of concentration and familiarity for each story on a scale of 0 to 1 after each run. The results showed that the concentration and familiarity ratings for the personal stories [0.78 ± 0.17 (mean ± SD) and 0.87 ± 0.16, respectively] were still higher than the common stories (0.69 ± 0.21 and 0.74 ± 0.19), indicating that personal stories elicited higher levels of concentration and familiarity. This might suggest that the inherent attention levels while reading personal versus common stories could differ, potentially introducing an unwanted confound between self-relevance and attention. Nevertheless, the concentration and familiarity ratings for common stories were also high enough for us to assume that the two types of stories were not qualitatively different. In future studies, however, it would be good to try using a single type of story with variable levels of self-relevance and valence. Second,

the prediction performances for decoding free-thinking and resting-state data were low and turned nonsignificant upon correction for multiple comparisons. Thus, they require careful interpretation with the risk of type I errors. Nevertheless, the significant predictions for self-relevance and valence within identical time windows might imply a reduced likelihood of these being false positives (Fig. 5*B*). Furthermore, significant valence predictions within the same time window across two independent datasets offer additional support for the generalizability of our predictive model (*SI Appendix*, Fig. S12). In addition, it may imply that the brain representations of spontaneous thought are highly complex and idiosyncratic. In a recent study, we showed that the brain representations of valence become increasingly idiosyncratic as thought topics become more self-relevant (6). Therefore, adopting a personalized modeling approach to studying spontaneous thought (e.g., extensive sampling of small-*N* design) could be considered in future studies. Also, we only tested the activation pattern-based models. Considering that a recent study convincingly showed that a functional connectivity-based model performed better than an activation-based model in predicting an emotional experience in naturalistic contexts (70), it is possible that functional connectivity-based models work better in our case as well. This should be tested in future studies. Third, the intrinsic interrelationship between self-relevance and valence might influence our results. As shown in *SI Appendix*, Fig. S3, there was a positive correlation between raw self-relevance and valence ratings, $r = 0.111$, $P < 2.220e\text{-}16$, potentially attributed to the well-documented self-positivity bias (71). To counteract the effects of this correlative nature between self-relevance and valence on our data and modeling, we implemented multiple strategies. These included constructing personal narratives on both positive (i.e., safety, pleasure) and negative topics (i.e., danger, pain) and conducting data quantization prior to predictive modeling. However, there were weak, but significant spatial correlations between the whole-brain or mPFC weight patterns of the two models (Fig. 2). In addition, the valence model appeared to be predictive of self-relevance at a time window near the reporting onset in Fig. 4*B*. Future studies should delve deeper into the nuanced relationship between the self-relevance and valence of spontaneous thoughts and their neural representations.

Despite these caveats, the current study introduces a unique approach to brain decoding of spontaneous thought—utilizing self-generated personal stories to develop brain-based predictive models of self-relevance and valence of spontaneous thought contents. As these models showed the potential to decode these spontaneous thought content dimensions during free-thinking or resting state, they hold the potential to address some basic science questions that are otherwise inaccessible. Overall, this study provides an important step toward developing brain models of internal thoughts and emotions during daydreaming.

Materials and Methods

Participants. Fifty-seven healthy right-handed participants completed the experiment. Participants provided written informed consent in compliance with the guidelines of the Sungkyunkwan University Bioethics Committee. The full study protocol was approved by the institutional review board. Participants with psychiatric, neurological, or systemic disorders and MRI contraindications were excluded. All participants had a normal or corrected-to-normal vision and were naïve to the purpose of the experiment. All participants were Koreans and spoke Korean as their first language. All experimental procedures were conducted using the Korean language. We included forty-nine participants [age = 22.8 ± 2.4 (mean ± SD), 21 female] in the final data given that we excluded eight participants total [six participants due to poor performance (e.g., did not focus on the task, slept during the scan, or did not fully understand the task), and two participants due to poor image quality].

Experimental Procedure. The experiment consisted of three components across sessions of two days (Fig. 1*B*): 1) an online interview on day 1, 2) an fMRI experiment with the story-reading and thought-sampling tasks on day 2, and 3) a post-scan survey on day 2. First, on day 1, after providing an overview of the experiment and having participants complete a pre-scan survey of self-report questionnaires, we conducted a one-on-one interview to create personal stories. After the interview, the participants read the common stories to match the level of familiarity between the personal and common story sets. Then, participants visited the lab again about 1 wk [7.32 ± 2.8 (mean ± SD) d] after their first visit for the fMRI experiment and post-scan survey. The fMRI experiment had a total of seven runs, which consisted of five story-reading runs (about 14 min per run) and two thought-sampling runs (about 6 min per run). We placed the thought-sampling runs in the first and last runs of the seven runs. We used MATLAB (MathWorks) and Psychtoolbox (version 3.0.16, http://psychtoolbox.org/) for stimuli presentation and behavioral data acquisition. After the fMRI experiment, we conducted a post-scan survey. For details of the interview, story-making procedure, story-reading and though-sampling tasks, and post-scan survey, please see *SI Appendix*, *Supplementary Methods*.

fMRI Data Acquisition and Analysis. Whole-brain MRI data were acquired on a 3T Siemens Prisma scanner at Sungkyunkwan University with a 64-channel head coil. High-resolution T1-weighted structural images were acquired with TR = 2,400 ms and TE = 2.34 ms. Functional echo-planar imaging (EPI) images were acquired with TR = 460 ms, TE = 27.2 ms, multiband acceleration factor = 8, field of view = 220 mm, 82 × 82 matrix, 2.7 × 2.7 × 2.7 mm^3 voxels, and 56 interleaved slices. The number of volumes was 812 for the thought-sampling runs and 1,855 for the story-reading runs. For details about the fMRI data preprocessing, general linear modeling, predictive modeling, model interpretation, and independent testing, please see *SI Appendix*, *Supplementary Methods*.

Data, Materials, and Software Availability. The data and codes used to generate the main figures, including the predictive models, are shared through Zenodo.org (https://zenodo.org/doi/10.5281/zenodo.10039368) (72). In-house Matlab codes for fMRI data analyses are available at https://github.com/canlab/CanlabCore (73) and https://github.com/cocoanlab/cocoanCORE (74).

ACKNOWLEDGMENTS. This work was supported by IBS-R015-D1 (Institute for Basic Science; to C.-W.W.) and 2021M3E5D2A01022515 (National Research Foundation of Korea; C.-W.W.). We thank Jinwon Park, Minie Jung, Hyebhin Yoon, and Sungwoo Lee for helping conduct the experiments and Jihoon Han for translating the experimental stimuli from Korean to English.

Author affiliations: [a]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon 16419, South Korea; [b]Department of Biomedical Engineering, Sungkyunkwan University, Suwon 16419, South Korea; [c]Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon 16419, South Korea; [d]Department of Psychological and Brain Sciences, Dartmouth College, NH 03755; and [e]Life-inspired Neural Network for Prediction and Optimization Research Group, Suwon 16419, South Korea

1. M. A. Killingsworth, D. T. Gilbert, A wandering mind is an unhappy mind. *Science* **330**, 932–932 (2010).
2. K. Christoff, Z. C. Irving, K. C. Fox, R. N. Spreng, J. R. Andrews-Hanna, Mind-wandering as spontaneous thought: A dynamic framework. *Nat. Rev. Neurosci.* **17**, 718–731 (2016).
3. J. Smallwood, J. W. Schooler, The science of mind wandering: Empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).
4. E. Klinger, W. M. Cox, Dimensions of thought flow in everyday life. *Imagin. Cogn. Pers.* **7**, 105–128 (1987).
5. E. Klinger, Goal Commitments and the content of thoughts and dreams: Basic principles. *Front. Psychol.* **4**, 415 (2013).
6. B. Kim, J. R. Andrews-Hanna, J. Han, E. Lee, C. W. Woo, When self comes to a wandering mind: Brain representations and dynamics of self-generated concepts in spontaneous thought. *Sci. Adv.* **8**, eabn8616 (2022).
7. J. R. Andrews-Hanna *et al.*, A penny for your thoughts: Dimensions of self-generated thought content and relationships with individual differences in emotional wellbeing. *Front. Psychol.* **4**, 900 (2013).

8. J. R. Andrews-Hanna *et al.*, The conceptual building blocks of everyday thought: Tracking the emergence and dynamics of ruminative and nonruminative thinking. *J. Exp. Psychol. Gen.* **151**, 628–642 (2022).
9. J. Smallwood, R. C. O'Connor, M. V. Sudbery, M. Obonsawin, Mind-wandering and dysphoria. *Cogn. Emotion* **21**, 816–842 (2007).
10. I. Marchetti, E. H. W. Koster, E. Klinger, L. B. Alloy, Spontaneous thought and vulnerability to mood disorders: The dark side of the wandering mind. *Clin. Psychol. Sci.* **4**, 835–857 (2016).
11. L. Kvavilashvili, A. Niedzwienska, S. J. Gilbert, I. Markostamou, Deficits in spontaneous cognition as an early marker of Alzheimer's disease. *Trends Cogn. Sci.* **24**, 285–301 (2020).
12. K. Christoff, Undirected thought: Neural determinants and correlates. *Brain Res.* **1428**, 51–59 (2012).
13. W. Heisenberg, C. Eckart, F. C. Hoyt, *The Physical Principles of the Quantum Theory* (The University of Chicago Science Series, The University of Chicago Press, Chicago, Ill, 1930), p. xii, 186 p.
14. A. Kucyi *et al.*, Prediction of stimulus-independent and task-unrelated thought from functional brain networks. *Nat. Commun.* **12**, 1793 (2021).
15. S.-M. Hung, P.-J. Hsieh, Mind wandering in sensory cortices. *Neuroimage Rep.* **2**, 100073 (2022).
16. A. Tusche, J. Smallwood, B. C. Bernhardt, T. Singer, Classifying the wandering mind: Revealing the affective content of thoughts during task-free rest periods. *Neuroimage* **97**, 107–116 (2014).
17. M. Sormaz *et al.*, Default mode network can support the level of detail in experience during active task states. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9318–9323 (2018).
18. B. Bellana, A. Mahabal, C. J. Honey, Narrative thinking lingers in spontaneous thought. *Nat. Commun.* **13**, 4585 (2022).
19. H. Lee, B. Bellana, J. Chen, What can narratives tell us about the neural bases of human memory? *Curr. Opin. Behav. Sci.* **32**, 111–119 (2020).
20. J. Xu, S. Kemeny, G. Park, C. Frattali, A. Braun, Language in context: Emergent features of word, sentence, and narrative comprehension. *Neuroimage* **25**, 1002–1015 (2005).
21. Y. Lerner, C. J. Honey, L. J. Silbert, U. Hasson, Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* **31**, 2906–2915 (2011).
22. Y. Yeshurun *et al.*, Same story, different story. *Psychol. Sci.* **28**, 307–319 (2017).
23. M. Nguyen, T. Vanderwal, U. Hasson, Shared understanding of narratives is correlated with shared neural responses. *Neuroimage* **184**, 161–170 (2019).
24. S. A. Nastase *et al.*, The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. *Sci. Data* **8**, 250 (2021).
25. J. Tang, A. LeBel, S. Jain, A. G. Huth, Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat. Neurosci.* **26**, 858–866 (2023).
26. P. Delamillieure *et al.*, The resting state questionnaire: An introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res. Bull.* **81**, 565–573 (2010).
27. J. Smallwood *et al.*, Self-reflection and the temporal focus of the wandering mind. *Conscious Cogn* **20**, 1120–1126 (2011).
28. J. R. Andrews-Hanna, J. S. Reidler, C. Huang, R. L. Buckner, Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* **104**, 322–335 (2010).
29. B. Baird, J. Smallwood, J. W. Schooler, Back to the future: Autobiographical planning and the functionality of mind-wandering. *Conscious Cogn.* **20**, 1604–1611 (2011).
30. D. Stawarczyk, S. Majerus, M. Maj, M. Van der Linden, A. D'Argembeau, Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychol. (Amst)* **136**, 370–381 (2011).
31. S. N. Cole, D. Berntsen, Do future thoughts reflect personal goals? Current concerns and mental time travel into the past and future. *Q. J. Exp. Psychol. (Hove)* **69**, 273–284 (2016).
32. D. Stawarczyk, H. Cassol, A. D'Argembeau, Phenomenology of future-oriented mind-wandering episodes. *Front. Psychol.* **4**, 425 (2013).
33. X. Song, X. Wang, Mind wandering in Chinese daily lives–an experience sampling study. *PLoS One* **7**, e44423 (2012).
34. K. C. R. Fox, K. Christoff, *The Oxford Handbook of Spontaneous Thought: Mind-wandering, Creativity, and Dreaming*, Oxford Library of Psychology (Oxford University Press, New York, NY, 2018), p. xvi, 611 pp.
35. B. Medea *et al.*, How do we decide what to do? Resting-state connectivity patterns and components of self-generated thought linked to the development of more concrete personal goals. *Exp. Brain Res.* **236**, 2469–2481 (2018).
36. J. N. Mildner, D. I. Tamir, Spontaneous thought as an unconstrained memory process. *Trends Neurosci.* **42**, 763–777 (2019).
37. P. Fossati *et al.*, In search of the emotional self: An fMRI study using positive and negative emotional words. *Am. J. Psychiatry* **160**, 1938–1945 (2003).
38. T. W. Schmitz, S. C. Johnson, Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neurosci. Biobehav. Rev.* **31**, 585–596 (2007).
39. K. J. Gorgolewski *et al.*, A correspondence between individual differences in the brain's intrinsic functional architecture and the content and form of self-generated thoughts. *PLoS One* **9**, e97176 (2014).
40. G. Northoff *et al.*, Self-referential processing in our brain–A meta-analysis of imaging studies on the self. *Neuroimage* **31**, 440–457 (2006).
41. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).
42. M. Corbetta, G. Patel, G. L. Shulman, The reorienting system of the human brain: From environment to theory of mind. *Neuron* **58**, 306–324 (2008).
43. A. Turnbull *et al.*, The ebb and flow of attention: Between-subject variation in intrinsic connectivity and cognition associated with the dynamics of ongoing experience. *Neuroimage* **185**, 286–299 (2019).
44. T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
45. R. A. Poldrack, G. Huckins, G. Varoquaux, Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry* **77**, 534–540 (2020).
46. G. Varoquaux *et al.*, Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage* **145**, 166–179 (2017).
47. L. Koban, P. J. Gianaros, H. Kober, T. D. Wager, The self in context: Brain systems linking mental and physical health. *Nat. Rev. Neurosci.* **22**, 309–322 (2021).
48. B. T. Denny, H. Kober, T. D. Wager, K. N. Ochsner, A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* **24**, 1742–1752 (2012).
49. M. Ceko, P. A. Kragel, C. W. Woo, M. Lopez-Sola, T. D. Wager, Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760–770 (2022).
50. J. Chikazoe, D. H. Lee, N. Kriegeskorte, A. K. Anderson, Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014).
51. L. Kohoutova *et al.*, Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat. Protoc.* **15**, 1399–1435 (2020).
52. G. H. Ngo *et al.*, Beyond consensus: Embracing heterogeneity in curated neuroimaging meta-analysis. *Neuroimage* **200**, 142–158 (2019).
53. M. Roy, D. Shohamy, T. D. Wager, Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16**, 147–156 (2012).
54. L. J. Chang, P. J. Gianaros, S. B. Manuck, A. Krishnan, T. D. Wager, A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* **13**, e1002180 (2015).
55. J. R. Andrews-Hanna, M. D. Grilli, Mapping the imaginative mind: Charting new paths forward. *Curr. Dir. Psychol. Sci.* **30**, 82–89 (2021).
56. V. Menon, 20 years of the default mode network: A review and synthesis. *Neuron* **111**, 2469–2487 (2023).
57. T. T. Raij, T. J. J. Riekki, Dorsomedial prefontal cortex supports spontaneous thinking per se. *Hum. Brain Mapp.* **38**, 3277–3288 (2017).
58. S. Aliko, J. Huang, F. Gheorghiu, S. Meliss, J. I. Skipper, A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Sci. Data* **7**, 347 (2020).
59. E. S. Finn, E. Glerean, U. Hasson, T. Vanderwal, Naturalistic imaging: The use of ecologically valid conditions to study brain function. *Neuroimage* **247**, 118776 (2022).
60. J. R. Binder *et al.*, Conceptual processing during the conscious resting state. A functional MRI study. *J. Cogn. Neurosci.* **11**, 80–95 (1999).
61. R. L. Buckner, M. E. Wheeler, The cognitive neuroscience of remembering. *Nat. Rev. Neurosci.* **2**, 624–634 (2001).
62. P. Qin, G. Northoff, How is our self related to midline regions and the default-mode network? *Neuroimage* **57**, 1221–1233 (2011).
63. C. G. Davey, J. Pujol, B. J. Harrison, Mapping the self in the brain's default mode network. *Neuroimage* **132**, 390–397 (2016).
64. X. Wang, B. Wang, Y. Bi, Close yet independent: Dissociation of social from valence and abstract semantic dimensions in the left anterior temporal lobe. *Hum. Brain Mapp.* **40**, 4759–4776 (2019).
65. J. R. Andrews-Hanna, J. S. Reidler, J. Sepulcre, R. Poulin, R. L. Buckner, Functional-anatomic fractionation of the brain's default network. *Neuron* **65**, 550–562 (2010).
66. U. Altmann, I. C. Bohrn, O. Lubrich, W. Menninghaus, A. M. Jacobs, The power of emotional valence-from cognitive to affective processes in reading. *Front. Hum. Neurosci.* **6**, 192 (2012).
67. F. Pereira *et al.*, Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* **9**, 963 (2018).
68. T. M. Mitchell *et al.*, Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
69. E. S. Finn, Is it time to put rest to rest? *Trends Cogn. Sci.* **25**, 1021–1032 (2021).
70. F. Z. Zhou *et al.*, Capturing dynamic fear experiences in naturalistic contexts: An ecologically valid fMRI signature integrating brain activation and connectivity. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.08.18.553808 (Accessed 10 September 2023).
71. S. E. Taylor, J. D. Brown, Illusion and well-being: A social psychological perspective on mental health. *Psychol. Bull.* **103**, 193–210 (1988).
72. H. Kim *et al.*, Brain Decoding of Spontaneous Thought: Predictive Modeling of Self-relevance and Valence Using Personal Narratives. Zenodo. https://zenodo.org/records/10039369. Deposited 25 October 2023.
73. T. D. Wager, (Cognitive and Affective Neuroscience Lab), CanlabCore. github. https://github.com/canlab/CanlabCore. Accessed 1 February 2018.
74. C.-W. Woo (Computational Cognitive Affective Neuroscience Laboratory), cocoanCORE. github. https://github.com/cocoanlab/cocoancore. Accessed 1 September 2018.